BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 6: e93668
doi: 10.3897/biss.6.93668

OPEN ACCESS

Conference Abstract

TaxonWorks as a Tool for Managing Large Biodiversity Projects

Dmitry A Dmitriev‡, Yanghui Cao‡, Christopher H. Dietrich§

‡ University of Illinois at Urbana-Champaign, Champaign, United States of America
§ Illinois Natural History Survey, Champaign, United States of America

Corresponding author: Dmitry A Dmitriev (arboridia@gmail.com)

Received: 18 Aug 2022 | Published: 23 Aug 2022

Citation: Dmitriev DA, Cao Y, Dietrich CH (2022) TaxonWorks as a Tool for Managing Large Biodiversity Projects. Biodiversity Information Science and Standards 6: e93668. https://doi.org/10.3897/biss.6.93668

Abstract

Large systematic revisionary projects incorporating data for hundreds or thousands of taxa require an integrative approach, with a strong biodiversity-informatics core for efficient data management to facilitate research on the group. Our original biodiversity informatics platform, 3i (Internet-accessible Interactive Identification) combined a customized MS Access database backend with ASP-based web interfaces to support revisionary syntheses of several large genera of leafhopers (Hemiptera: Auchenorrhyncha: Cicadellidae). More recently, for our National Science Foundation sponsored project, "GoLife: Collaborative Research: Integrative genealogy, ecology and phenomics of deltocephaline leafhoppers (Hemiptera: Cicadellidae), and their microbial associates", we selected the new open-source platform TaxonWorks as the cyberinfrastructure. In the scope of the project, the original "3i World Auchenorrhyncha Database" was imported into TaxonWorks. At the present time, TaxonWorks has many tools to automatically import nomenclature, citations, and specimen based collection data. At the time of the initial migration of the 3i database, many of those tools were still under development, and complexity of the data in the database required a custom migration script, which is still probably the most efficient solution for importing datasets with long development history.

At the moment, the World Auchenorrhyncha Database comprehensively covers nomenclature of the group and includes data on 70 valid families, 6,816 valid genera, 47,064 valid species as well as synonymy and subsequent combinations (Fig. 1). In addition, many taxon records include the original citation, bibliography, type information,

© Dmitriev D et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

etymology, etc. The bibliography of the group includes 37,579 sources, about 1/3 of which are associated with PDF files. Species have distribution records, either derived from individual specimens or as country and state level asserted distribution, as well as biological associations indicating host plants, predators, and parasitoids.

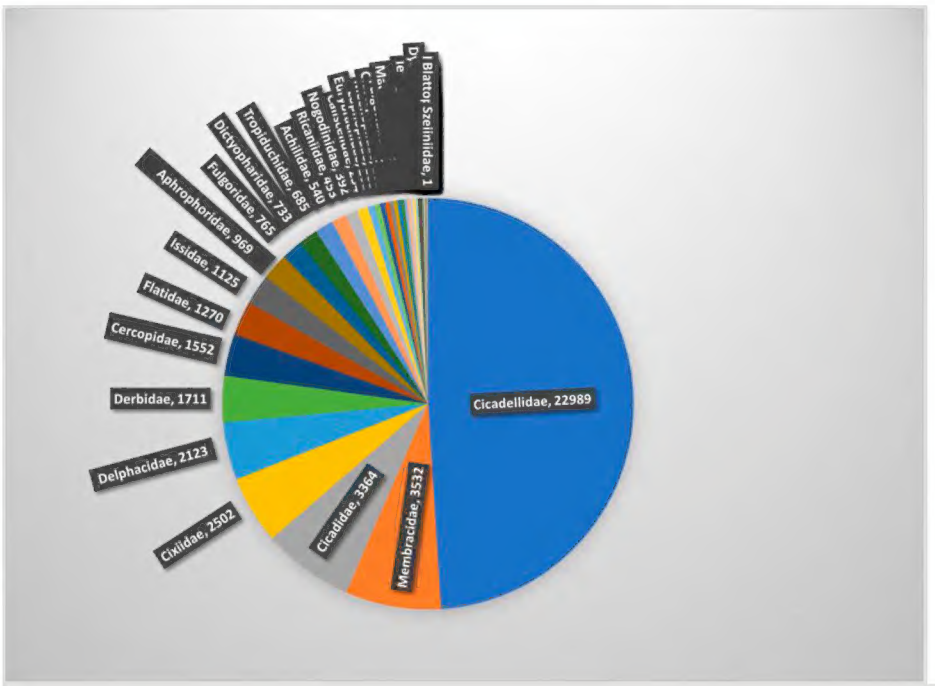

Figure 1.

Diversity of Auchenorrhyncha. Number of species in each family.

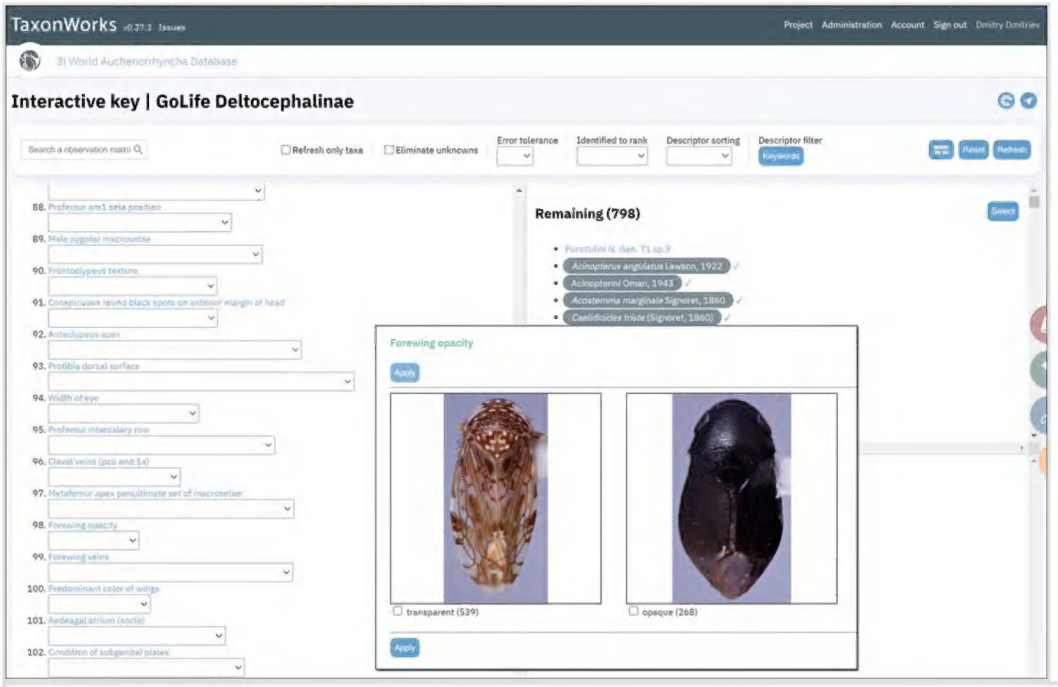

Figure 2.

An interactive key to the subfamily Deltocephalinae.

Observation matrices in TaxonWorks are designed to handle morphological data associated with taxa or specimens. The matrices may be used to automatically generate interactive identification keys and taxon descriptions. They can also be downloaded to be imported, for example, into Lucid builder, or to perform phylogenetic analysis using an external application. At the moment there are 36 matrices associated with the project. The observation matrix from GoLife project covers 798 taxa by 210 descriptors (most of which

are qualitative multi-state morphological descriptors) (Fig. 2). Illustrations are provided for 9,886 taxa and organized in the specialized image matrix and could be used as a pictorial key for determination of species and taxa of a higher rank.

For the phylogenetic analysis, a dataset was constructed for 730 terminal taxa and >160,000 nucleotide positions obtained using anchored hybrid enrichment of genomic DNA for a sample of leafhoppers from the subfamily Deltocephalinae and outgroups. The probe kit targets leafhopper genes, as well as some bacterial genes (endosymbionts and plant pathogens transmitted by leafhoppers). The maximum likelihood analyses of concatenated nucleotide and amino acid sequences as well as coalescent gene tree analysis yielded well-resolved phylogenetic trees (Cao et al. 2022). Raw sequence data have been uploaded to the Sequence Read Archive on GenBank. Occurrence and morphological data, as well as diagnostic images, for voucher specimens have been incorporated into TaxonWorks.

Data in TaxonWorks could be exported in raw format, get accessed via Application Programming Interface (API), or be shared with external data aggregators like Catalogue of Life, GBIF, iDigBio.

Keywords

Hemiptera, Homoptera, Auchenorrhyncha, Cicadellidae, Deltocephalinae, leafhoppers, databases, checklist, taxonomy, systematics

Presenting author

Dmitry A Dmitriev

Acknowledgements

Many people have contributed to the design and development of TaxonWorks, the detailed information is summarized here: https://github.com/SpeciesFileGroup/taxonworks/blob/development/CONTRIBUTORS.md. The work on 3i World Auchenorrhyncha Database is supported by the Species File Group and the U.S. National Science Foundation (NSF DEB 16-39601)

Funding program

U.S. National Science Foundation (NSF DEB 16-39601)

Grant title

GoLife: Collaborative Research: Integrative genealogy, ecology and phenomics of deltocephaline leafhoppers (Hemiptera: Cicadellidae), and their microbial associates

References

- Cao Y, Dietrich C, Zahniser J, Dmitriev D (2022) Datasets for "Dense sampling of taxa and characters improves phylogenetic resolution among deltocephaline leafhoppers (Hemiptera: Cicadellidae: Deltocephalinae)". University of Illinois at Urbana-Champaign https://doi.org/10.13012/b2idb-8842653_v1